Soil specific surface area determination by visible near-infrared spectroscopy

Knadel, Maria; Arthur, Emmanuel; Weber, Peter; Moldrup, Per; Greve, Mogens Humlekrog; Chrysodonta, Zampella Pittaki; De Jonge, Lis W.

Published in:
Soil Science Society of America Journal

DOI (link to publication from Publisher):
10.2136/sssaj2018.03.0093

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Published September 20, 2018

Soil Specific Surface Area Determination by Visible Near-Infrared Spectroscopy

Maria Knadel*
Emmanuel Arthur
Peter Weber
Aarhus Univ.
Dep. of Agroecology
Blichers Allé, Tjele, 8830-Denmark

Per Moldrup
Aalborg Univ.
Dep. of Civil Engineering
Thomas Manns Vej 23
Aalborg, 9200-Denmark

Mogens Humlekrog Greve
Zampella Pittaki Chrysodonta
Lis W. de Jonge
Aarhus Univ.
Dep. of Agroecology
Blichers Allé, Tjele, 8830-Denmark

The soil specific surface area (SSA) affects soil physical and chemical properties. Numerous studies applied visible near-infrared spectroscopy (Vis–NIRS) to estimate clay content (particles < 2 μm). Since SSA is better defined and more directly related to particle size distribution and mineralogy than clay content, predictions of SSA from Vis–NIRS are expected to be better than that for clay. Thus, the aims of this study were to (i) test the feasibility of using Vis–NIRS for SSA determination, (ii) compare the predictive ability of Partial Least Squares (PLS) model of SSA with that of clay, (iii) identify important wavelengths using interval Partial Least Squares (iPLS) regression, and to test if the application of iPLS improves the predictive ability of the models. A total of 550 soil samples with a wide range in SSA (3–437 m^2 g^{-1}) and clay content (1–83%) was divided into a calibration and a validation set. The PLS models had similar predictive ability for SSA (ratio of performance to interquartile range, RPIQ = 1.7) and clay content (RPIQ = 1.6). Utilizing iPLS led to only limited improvement in the prediction accuracy (RPIQ of 1.8 and 1.7 for SSA and clay content, respectively), yet decreased the number of relevant wavelengths and indicated a higher specificity of SSA over the broader spectral response of clay. The important wavelengths for SSA and clay predictions were indicative of the organo-mineral content and its interactions, including spectral response from not only iron oxides and minerals but also organic matter due to masking effect of the non-complexed organic carbon on the mineral phases of some of the soils.

Abbreviations: EGME, ethylene glycol monoethyl ether; iPLS, interval partial least squares; OM, organic matter; PC, principal component; PCA, principal component analysis; PLS, partial least squares; RPIQ, ratio of performance to interquartile range; RMSEP, root mean squared error of prediction; SOC, soil organic carbon; SSA, soil specific surface area; vis-NIRS, visible near-infrared spectroscopy.

Core Ideas

- **Feasibility of using vis–NIRS for SSA determination in a highly variable data set was investigated**
- **Acceptable predictive ability for SSA using PLS and iPLS was obtained.**
- **iPLS resulted in little improved accuracy for both SSA and clay**
- **iPLS indicated a higher specificity of SSA than of clay response**
- **The important wavelengths reflected the organo-mineral composition and its interactions**

The soil specific surface area (SSA) is a basic soil property and indicates the surface area per unit mass of soil. It affects physical and chemical properties of the soil and is important for processes such as water retention and movement, ion exchange reactions, contaminant adsorption, microbial attachment, nutrient dynamics, soil aggregation, and irrigation management (Pennell, 2002). The SSA can vary greatly depending on the soil type and the method used for its determination. Different soils present differences in mineralogical and organic composition as well as in particle-size distribution resulting in large variations in the amount of available reactive surfaces. For instance, soils rich in clay and organic matter can have SSA of up to 8×10^5 m^2 kg^{-1}, whereas the SSA for sandy soils can be as low as 1×10^3 m^2 kg^{-1} (Pennell, 2002).

Due to the importance of SSA, there is a worldwide interest and need for its accurate determination. Although SSA is measured routinely by scientists, agronomists, and engineers around the world, there is no common standard method available for its determination. The existing methods measure different surfaces of the

Soil Sci. Soc. Am. J. 82:1046–1056
doi:10.2136/sssaj2018.03.0093
Received 2 Mar. 2018.
Accepted 5 June 2018.
*Corresponding author (maria.knadel@agro.au.dk).

soil leading to different SSA estimates for a given sample. The SSA can be directly estimated by a measure of the particles shape and dimension (Borkovec et al., 1993). It can also be indirectly determined by sorption of gases or polar liquids. The adsorption of gases on solid surfaces with N_2 and CO_2 is used in combination with the Brunauer–Emmett–Teller (BET) equation (Kim et al., 2016). Due to the fact that these gases do not penetrate into the interlayer surfaces of clay minerals, they provide a measure of only the external specific area of a soil (Pennell, 2002). The other method consists of the measurement of polar liquids or molecules from solution to the surface (Heister, 2014); these polar liquids include: water (Tuller and Or, 2005; Arthur et al., 2018) ethylene glycol and ethylene glycol monoethyl ether (EGME) (Cerato and Lutenegger, 2002), and methylene blue (Hang and Brindley, 1970). Application of polar molecules measures the total SSA (internal and external) as they can penetrate the interlayers of clay minerals (Pennell, 2002). Among the above-mentioned methods, sorption of EGME is the most commonly used. There is, however, many disadvantages associated with its usage, such as long measurement times, complications associated with the monolayer coverage, effect of exchangeable cation species, capillary condensation, and environmental concerns about its disposal (Hang and Brindley, 1970; Tiller and Smith, 1990; Quirk and Murray, 1999; Pennell, 2002). Thus, a rapid, repeatable alternative to the above-mentioned methods to estimate SSA would be advantageous.

Visible near-infrared spectroscopy (Vis–NIRS) is a fast and reproducible method for soil analysis that could provide such an alternative. This method does not involve chemicals, uses small sample amounts that can be reused, involves minimal sample preparation (air drying and sieving). Moreover, a single spectrum of the soil can be used for the determination of multiple soil properties such as the soil textural fractions and organic matter content (Stenberg et al., 2010). Vis–NIRS covers the spectral range between 400 and 2500 nm. This region is dominated by overtones and combinations of the fundamental vibrations from the mid-infrared region (2500–25000nm). Most of the absorptions in this region are characteristic of clay and organic matter and are dominated by C–H, O–H, N–H and metal–OH bonds (Hunt, 1977; Clark et al., 1990). During the past decades, Vis–NIRS-based spectral models to evaluate soil physical, biological and chemical properties have been published (Stenberg et al., 2010; Soriano-Disla et al., 2013). The application of Vis–NIRS to more advanced soil parameters such as the degree of organic matter coverage of fine mineral particles (Hermansen et al., 2016), soil densities (Katuwal et al., 2017), soil binding capacity (Paradelo et al., 2016), and water repellency have been also reported (Kim et al., 2014; Knadel et al., 2016). Numerous studies have successfully applied Vis–NIRS to determine clay content (Soriano-Disla et al., 2013), a property strongly related to SSA. Despite this great potential of vis – NIR, only few studies have applied it for soil SSA estimation. In two studies on some Israeli soils, SSA was estimated with the EGME as the reference method and predicted with NIRS (Ben-Dor and Banin, 1995) and Vis–NIRS (Ben-Dor et al., 2008). In the former work, conducted on air dried soils and with SSA values ranging from 10.8 to 369 $m^2\ g^{-1}$, root mean squared error of prediction (RMSEP) of 50.2 $m^2\ g^{-1}$ and R^2 of 0.69 were reported (Ben-Dor and Banin, 1995). In the latter study, higher predictive ability of the SSA model was achieved (RMSEP of 20.9 $m^2\ g^{-1}$ and R^2 = 0.92) even though the analysis was based on field moist soils (SSA range: 27.3–407.1 $m^2\ g^{-1}$) (Ben-Dor et al., 2008). Ben-Dor and Banin (1995) reported that clay content and SSA were related to OH in water but also to metal-OH in the mineral lattice of the minerals. The studies mentioned above are relatively old and explored the possibility of predicting SSA from spectroscopy, but utilized data from one region only. For widely applicable soil property estimation, it is useful to encompass samples from several geographical regions (hence, different soil types). The SSA, despite being difficult to measure, is well-defined compared with clay content, which is delineated at an arbitrary particle size of 2 μm. Second, SSA is more intimately linked to soil properties such as clay mineralogy and cation exchange capacity, and is thus expected to be more strongly expressed in a Vis–NIRS spectra compared with clay content. Based on the above-mentioned considerations, the objectives of the current study were to: i) test the feasibility of using Vis–NIRS for SSA determination for a large data set with a wide range of soils from different geographical areas, (ii) compare the predictive ability of the Vis–NIRS model for predicting SSA with that for predicting clay, (iii) identify important variables for SSA and clay predictions, and to test if additional variables selection improves the predictive ability of the calibration models.

MATERIALS AND METHODS

Soil Samples

The investigated dataset included 550 soil samples from different parts of the world. The samples originated from five geographic regions including: three countries in North America (Cuba, Nicaragua, United States; N = 68), four countries in South America (Brazil, Colombia, Ecuador, Peru; N = 10), five countries in Africa (Ghana, Nigeria, Mozambique, South Africa, Zimbabwe; N = 11), three countries in Asia and Oceania (India, New Zealand, Sri Lanka; N = 9), and three countries in Europe (Denmark, Germany, Norway, N = 452). The soil samples were all air dried and sieved to 2 mm prior to laboratory analyses.

Measurement of Soil Properties

Soil texture was determined by a combination of wet-sieving and pipette or hydrometer methods (Gee and Or, 2002) after removal of organic matter and carbonates. The soil organic carbon (SOC) content was obtained either by the oxidation of carbon at 1800 °C with an organic elemental analyzer coupled to a thermal conductivity detector (Thermo Fisher Scientific Inc., Waltham, MA) or by wet combustion following the Walkley–Black method (Nelson and Sommers, 1996). The reference soil specific surface area (SSA) was estimated by the retention of ethylene glycol monoethyl ether (EGME) at monolayer coverage (Pennell, 2002).

The soil samples were not pretreated (no removal of organic carbon nor ion saturation) prior to the measurements.

Spectral Measurements

The spectral data was obtained by scanning the samples with a Vis–NIRS sensor (NIRS DS2500, FOSS, Hillerød, Denmark) in a temperature and humidity controlled NIRS laboratory (temperature of 23°C, humidity of 48%). The measurements were done in the spectral range between 400 and 2500 nm and a spectral interval of 0.5 nm. The scanning procedure was based on an internally developed scanning protocol described briefly below. Approximately 50 g of each soil sample were loaded to a sample holder with a quartz window and measured in seven spots. The collected seven scans were averaged to one scan for each sample. The output soil spectra were recorded as absorbance, Abs = [log(1/R)], where R is reflectance, and absorbance spectra were used further in the modeling phase.

Multivariate Data Analysis

Before the development of spectral calibration models, the entire dataset was subdivided into calibration and validation data sets. This ensured creation of robust calibration models that were validated with the data set not involved in the calibration step. First, Principal Component Analysis (PCA) was performed on the entire data set to compress the data into a smaller number of latent variables (principal components; PCs) (Webster, 2001). Further, the Kennard–Stone algorithm representing a sequential type of function (Kennard and Stone, 1969) was applied to the scores of the first three PCs. It enabled a selection of a calibration data set that had a uniform distribution over the spectral space. The function is based on Euclidean distance. The first two selected samples represented the most extreme spectra within the entire data set. Then the third sample was chosen so that it was the furthest away from the two already selected. This was continued until the chosen number of samples has been selected. In this case two thirds of the entire data set was chosen for calibration (440 samples) and the remaining one-third (110 samples) was left out for validation purposes.

Calibration and Validation Models

Before development of the models, different preprocessing techniques were applied to spectral data and included scatter corrections (Multiplicative Scatter Correction and Standard Normal Variate) and derivatives (first and second Savitzky-Golay) (Savitzky and Golay, 1964; Barnes et al., 1989; Martens and Næs, 1989; Barnes et al., 2003).

First, Partial Least Squares (PLS) regression with the nonlinear iterative partial least square algorithm (Nørgaard et al., 2000) was used to generate Vis–NIRS calibration models for SSA and clay content. The important wavelengths explaining the variability in SSA and clay, as indicated by the regression coefficients, were listed and assigned to the possible components. Further, Interval Partial Least Squares (iPLS) regression was tested. The iPLS method selects a relevant subset of intervals (a subset of adjacent wavelengths) for each of the soil properties individually, which results in superior models in comparison with using the entire vis – NIR spectrum, thus decreasing the models' complexity and removing irrelevant information from the spectra (Nørgaard et al., 2000). The iPLS method was used in the forward mode meaning that the additional intervals of spectra were successively included in the analysis in search for the best combination of variables. As reported by Hermansen et al. (2017), the interval size affects the accuracy of the models, thus the following spectral intervals sizes were additionally tested (25, 50, 100, and 150). Calibration models using 440 samples and both PLS and iPLS were developed on untreated and pre-processed spectra with the full-cross validation method (leave-one-out) and validated independently with 110 samples. All multivariate data analysis was done in Matlab software using PLS toolbox 8.2 (Eigenvector Research Inc., Manson, WA).

The best spectral preprocessing as well as the optimal spectral interval size for cross-validation was considered to be the one resulting in models with the lowest root mean squared error of cross-validation (RMSECV) and the highest Pearson R^2. The optimal number of factors (NF) was selected as the point after which RMSECV did not decrease significantly (Gowen et al., 2011). Validation models were evaluated further with the RMSEP, R^2 and the ratio of RMSEP to interquartile range (Q3-Q1), RPIQ (Bellon-Maurel et al., 2010).

RESULTS AND DISCUSSION

Soil Samples

The investigated data set covered a wide range of soil textures and included clayey (clay content > 80%) to extremely sandy (sand content > 90%) soils (Fig. 1a). The SSA and clay content for the entire data set ranged from 3 to 437 $m^2\ g^{-1}$ and 1 to 83%, respectively (Table 1).

The SSA is, among other soil properties, highly related to the organo-mineral content and its interactions. For instance, the degree of saturation of clay minerals by organic matter induce strong changes in their surface properties (Kaiser and Guggenberger, 2003). Therefore, the high variability in SOC, clay, and clay mineralogy can result in a different behavior of the investigated soil samples in respect to e.g., potential for organic matter coating of clay minerals, hence affecting SSA. This behavior was explored using the relationship between clay and SOC contents for the five geographic regions (Fig. 1b). As reported by Dexter et al. (2008) clay has a limited capacity for SOC complexation, which was defined as a ratio (n) of clay to SOC of 10, and was illustrated here with an n saturation line (Fig. 1b). Above or below this threshold, a surplus of clay or SOC in the form of non-complexed clay or non-complexed SOC is present, respectively. According to Fig. 1b, the subset of samples with n below 10 was represented mainly by the European soils whereas, nearly all the soils from the remaining four regions (and some European soils) exhibited n above 10. This clear separation indicates that the samples from the Americas, Africa, Asia and Oceania were unsaturated with organic matter (SOC in the complexed form), and majority of European samples had the highest potential for coating with organic matter (SOC in the non-complexed form)

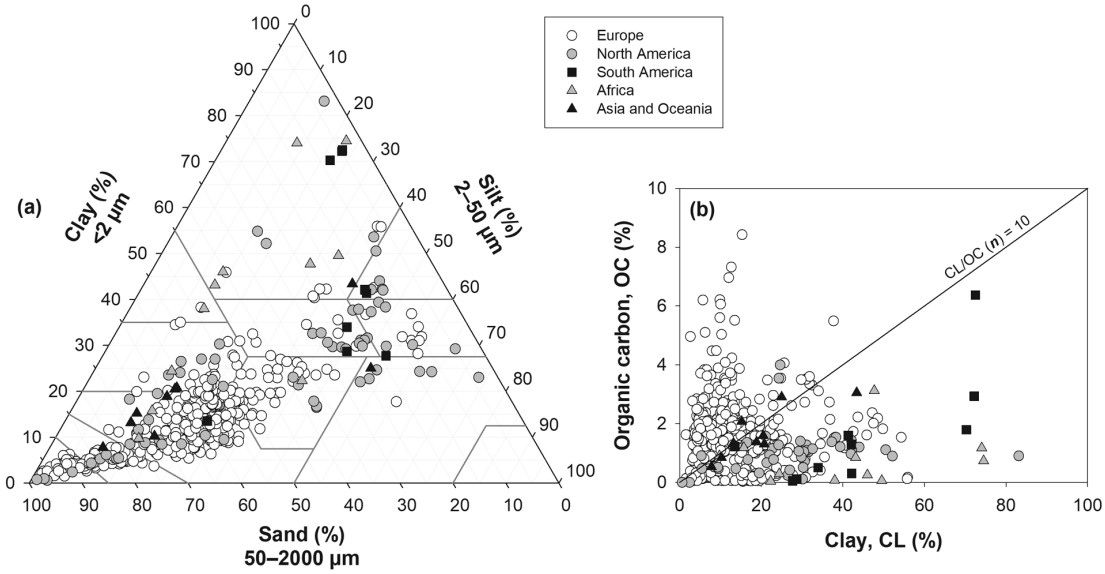

Fig. 1. a) Distribution of the soil samples in the USDA soil texture triangle and b) distribution of the soil samples around the *n* = 10 line (i.e., clay/organic carbon [OC] = 10) when plotting OC as a function of clay (*N* = 550).

that can potentially affect the SSA (Kaiser and Zech, 1998). Therefore, not only the differences in clay and its mineralogy but also the form (complexation status) of SOC is expected to have an effect on further SSA predictions.

The correlation between the SSA and the remaining soil properties for the entire data set was the highest with clay (r = 0.80), whereas with silt, sand and SOC it was 0.35, −0.66 and −0.03, respectively. Moreover, a simple multiple linear regression using clay, silt, sand and SOC as predictors for SSA of the entire data set was developed. Sand was excluded from the regression due to high collinearity with clay and silt, whereas, silt and SOC did not contribute significantly (P = 0.403 and P = 0.565, respectively) resulting in a simple equation with clay as a driving factor for SSA determination (SSA = −0.496 + 2.620 × clay) (Fig. 2). However, knowing that clear differences as to the importance of clay and SOC for SSA for the soils with n values below and above 10 can be expected (Fig. 1b), multiple linear regressions for the two subsets were additionally developed. For soils with n < 10 both clay and SOC were the significant predictors (P < 0.001), with SOC being positively correlated to SSA (SSA = −7.333 + 2.550 × clay + 3.790 × SOC). Similarly for the soils with n > 10, clay and SOC were significantly contributing (P < 0.001 and P = 0.001, respectively), nonetheless, this time SOC was correlated negatively to SSA (SSA = 1.690 + 2.922 × clay– 9.931 × SOC). The antagonistic roles played by organic matter are attributed to different relationships between the amount of organic matter, its types, differences in mineralogy and porosity (Petersen et al., 1996; Kaiser and Guggenberger, 2003; Bartoli et al., 2007; Arthur et al., 2013; Ding et al., 2013). If the non-complexed SOC is mainly sorbed to the surface of clay minerals the SSA increases with the increase of the non-complexed SOC, hence positive correlation. Conversely, a negative correlation between SOC and SSA was observed for the samples with the complexed SOC as a predominant form. Arthur et al. (2013) recognized also differences in SSA in respect to n values and concluded that for soils with high SOC values (n < 10) a part of SSA was masked due to organic matter sorption onto the clay minerals, resulting in the reduction of the total surface area. Moreover, the amount of clay content itself can lead to the different relationships between SOC and SSA. If the soils present a low clay and a high SOC contents, SOC is expected to coat the mineral grains and increase the SSA. Whereas, if the soils present a high clay content and thus exhibit much higher surface area, the SOC coating of the mineral surfaces can either reduce the SSA or has no impact.

Soil Spectral Signatures

The influence of clay content and mineralogy on the spectra is expected to be a valuable feature for predicting SSA from the vis – NIR spectra. Figure 3 illustrates differences in the spectral signatures among three selected soil samples including a smectitic,

Table 1. General statistics for the entire (*N* = 550), calibration (*N* = 440) and the validation (*N* = 110) data sets.

Property†‡	Mean	Max	Min	Range	SD	Variance	Q1	Q3
SSA, $m^2\ g^{-1}$	42 (42, 43)	437 (437, 116)	3 (3, 4)	434 (434, 111)	39 (42, 27)	1551 (1758, 34)	19 (16, 24)	50 (50, 49)
Clay, %	16 (16, 16)	83 (83, 50)	1 (1, 3)	82 (82, 46)	12 (13, 9)	143 (157, 86)	8 (8, 11)	20 (21, 19)
Silt, %	26 (25, 29)	73 (73, 58)	1 (1, 3)	72 (72, 55)	13 (14, 10)	179 (194, 106)	15 (14, 11)	32 (32, 34)
Sand, %	58 (59, 55)	98 (98, 94)	2 (2, 11)	96 (96, 83)	22 (22, 18)	465 (499, 316)	49 (50, 49)	71 (74, 63)
SOC, %	1.58 (1.59, 1.55)	8.42 (8.42, 5.6)	0 (0, 0.03)	8.42 (8.42, 5.57)	1.16 (1.23, 0.86)	1.35 (1.51, 0.74)	0.93 (0.88, 1.14)	2.01 (2.01, 1.95)

† Q1, the first quartile; Q3, the third quartile; SD, standard deviation; SOC, soil organic carbon; SSA, soil specific surface area.

‡ The first value is for the entire data set, the first value in brackets is for the calibration data set and the second value in the brackets is for the validation data set.

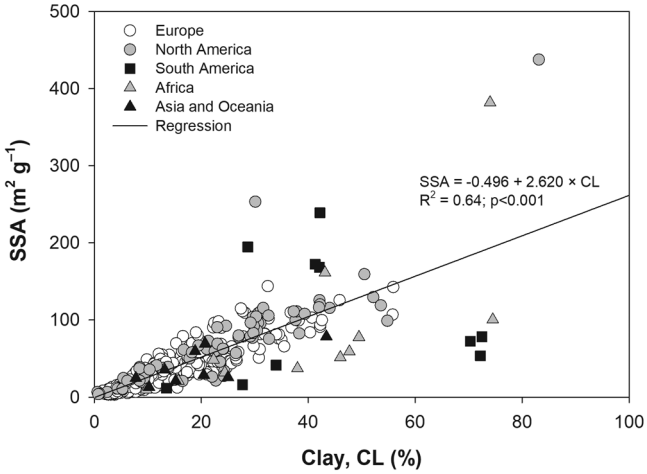

Fig. 2. Linear regression relation between specific surface area (SSA) and clay content for the entire data set.

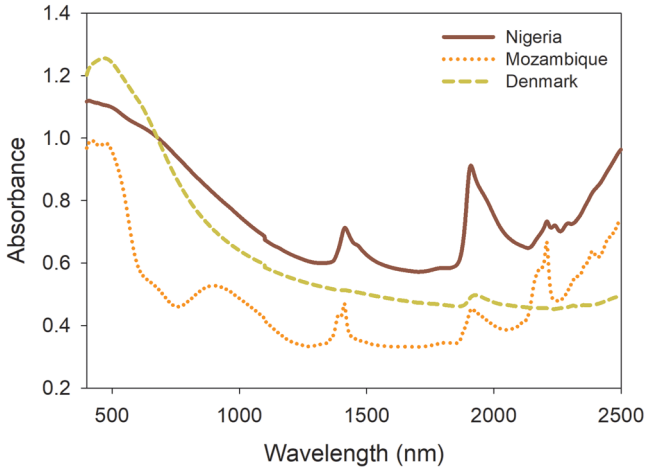

Fig. 3. Vis–NIRS spectra of three selected soil samples, representing a smectitic (Nigeria), a kaolinitic (Mozambique), and an illitic sample (Denmark).

kaolinitic and illitic soil. Smectites are known to be the most active of clay minerals due to the largest specific surface area and electrochemical activity (Ben-Dor, 2002). The selected smectitic sample originating from Nigeria exhibits a large SSA values (437 $m^2\ g^{-1}$; Fig. 3) and distinct spectral features. Smectitic soils present specific and very clear absorption features near 1400, 1900, and 2000 nm (Bishop et al., 1994). These were very pronounced in the Nigerian soil sample, which also had the highest clay content (83%). The peak near 1400 nm can be assigned to the first overtone of O-H stretch in its octahedral layer, while the peak near 1900 m originates from the OH bond due to combination vibrations of water bound in the interlayer lattice. Small shoulders near 1470 and 1970 nm are due to adsorbed water onto the mineral surface (Bishop et al., 1994). The soil sample from Mozambique represents a kaolinitic soil with SSA of 51 $m^2\ g^{-1}$. It exhibited very pronounced absorption peaks near 1400 and 1900 nm related to OH bonds (Post and Noble, 1993; Clark, 1999). Due to elevated clay content (46%), pronounced doublets near 1400 and 2200 nm occurred. The peak near 2200 nm was a result of a combination of Al-OH bend and O-H stretch (Bishop et al., 1994). Moreover, the Mozambican sample showed distinct absorption values in the visible range with peaks at 420, 470, and 920 nm attributed to iron oxides (goethite) (Scheinost, 1998). The Danish soil exhibited the lowest SSA (3 $m^2\ g^{-1}$) and the lowest clay content (4%) of the three samples in Fig. 3. Danish soils are dominated by illites, which are the typically formed minerals in colder regions (Moberg, 1990; Ben-Dor, 2002). The most common absorption bands for illites are present at 1400, 1900, and 2200 nm (Stenberg et al., 2010). However, due to the very low clay content only the band near 1900 nm related to OH was observable. The absorptions near 1400 and 2200 nm were very weak and barely visible. Moreover, the Danish sample had the highest SOC content (1.5%) leading to highest values of absorbance in the visible region, with a typical convex alike shape of the spectrum related to the presence of organic matter (Ben-Dor, 2002; Viscarra Rossel and Hicks, 2015).

As showed in the literature particle size affects the vis–NIR spectra. The finer the soil the higher overall reflectance (Stenberg et al., 2010). Sadeghi et al. (2018) documented it for six different soils from Arizona separated into seven particle/aggregate size classes. Particle size is intimately linked to clay content and SSA and is therefore important. This effect, however, was not present here as the soils were air dried and sieved to 2 mm (constant particle size). Subsequently, any observed differences in the spectra are due to the differences in the clay mineralogy, OM and water content.

Principal Component Analysis and data subdivision

Principal component analysis was performed on the entire spectral data set (550 samples). As the first three PCs accounted for a total of 98% of the variation in the spectral data Kennard-Stone algorithm was applied to the first three PCs only. Considering the variability in spectral data, the Kennard-Stone algorithm assured a representative selection of the calibration data set reflecting the variation in the mineral and organic matter of the investigated soils. Thus, the calibration data set showed some bias and covered the extreme and mean values of both SSA (4–437 $m^2\ g^{-1}$) and clay content (3–50%) in the data set (Table 1). This should facilitate more global yet, most likely less accurate models.

Partial Least Square (PLS) and interval Partial Last Squares (iPLS) Regression for SSA Determination

The results for PLS and iPLS for SSA determination showed clear improvement in the PLS calibration performance for SSA for the applied spectral preprocessing methods over untreated spectra (Table 2). The best PLS calibration model was achieved using the first derivative spectra and resulted in RMSECV = 21.4 $m^2\ g^{-1}$ and R^2 = 0.74, with eight factors (Fig. 4a). Similarly, better performance of the iPLS models for SSA was obtained after preprocessing the spectral data (Table 2, Fig. 4c). Although there seemed to be an effect of interval number on the calibration performance, there was no clear trend among the different preprocessing techniques. Nevertheless, the best model when iPLS was applied was likewise obtained on first derivative spectra (RMSECV = 19.6 $m^2\ g^{-1}$ and R^2 = 0.78, NF = 9) (Fig. 4a), using 50 nm intervals accounting for

Table 2. Partial Least Squares regression (PLS) and interval PLS (iPLS) cross-validation results for soil specific surface area. †‡

Spectra pretreatment	Method	IN (nm)	NV	RMSECV, $m^2\ g^{-1}$	R^2	Bias	NF
No pretreatment (Abs.)	PLS	–	2100	26.6	0.60	0.01	4
	iPLS	150	750	25.3	0.63	–0.04	4
		100	700	26.3	0.61	–0.07	4
		50	100	25.9	0.62	–0.02	5
		25	75	24.7	0.65	–0.06	5
SNV	PLS	–	2100	23.2	0.69	–0.03	5
	iPLS	150	450	23.3	0.69	–0.10	6
		100	600	22.6	0.71	–0.06	6
		50	400	22.7	0.70	–0.19	6
		25	175	23.22	0.69	–0.11	6
MSC	PLS	–	2100	22.6	0.71	–0.07	8
	iPLS	150	1050	20.5	0.76	–0.13	8
		100	800	21.1	0.75	–0.13	7
		50	450	21.1	0.75	–0.26	8
		25	250	22.9	0.70	–0.21	6
first der.	PLS	–	*2100*	*21.4*	*0.74*	*–0.16*	*8*
	iPLS	150	300	24.1	0.67	–0.04	5
		100	900	24.6	0.66	–0.09	6
		50	*1000*	*19.6*	*0.78*	*–0.27*	*9*
		25	100	21.1	0.75	0.07	10
second der.	PLS	–	2100	22.3	0.72	–0.22	6
	iPLS	150	1050	22.4	0.71	–0.27	5
		100	100	22.1	0.72	0.29	9
		50	150	21.5	0.74	–0.09	10
		25	125	22.6	0.72	–0.14	8

† Abs., absorbance; first der. and second der., the first and second Savitzky-Golay derivatives; IN, number of intervals; iPLS- interval Partial Least Squares regression; MSC, multiplicative scatter correction; NF, number of factors; NV, number of variables; PLS, Partial Least Squares regression; RMSECV, Root Mean Square Error of Cross-Validation; SNV, standard normal variate.

‡ Numbers in italics denote the best models.

700 variables. The application of iPLS not only resulted in a lower RMSECV and a higher coefficient of determination but also in the use of much fewer variables as opposed to using PLS on the entire vis – NIR spectrum (2100 variables).

Partial Least Square (PLS) and Interval Partial Last Squares (iPLS) Regression Results for Clay

The evaluation statistics for the clay PLS and iPLS calibration models are reported in Table 3. Slightly better performance of the PLS calibration models can be observed for the models based on preprocessed spectra than for the model based on the absorbance spectra. The best PLS model was obtained on second derivative spectra with RMSECV = 7.6%, R^2 = 0.63, and using seven factors (Fig. 4b). For clay also, the application of the iPLS method was advantageous as it decreased the number of variables in the best iPLS clay model to 1050, and resulted in a better model performance (RMSECV = 6.2%; R^2 = 0.75) when comparing it with the PLS method (Fig. 4d). The number of factors, however, increased in the iPLS (NF = 10) compared with the best PLS model (NF = 7).

Regression Coefficients

As SSA is a better defined soil property than clay, it is important to understand which spectral regions were the significant variables in the developed models. Figure 5a and 5b illustrate regression coefficients from the best SSA (based on the first derivative spectra and 7 factors) and clay (based on the second derivative and 8 factors) PLS calibration models, respectively, versus the wavelengths. Large regression coefficients indicate that the wavelengths play an important role in the determination of the response variables (SSA and clay). It was not possible to assign all of the important wavelengths to the potential chemical components, because the vibrations present in the vis – NIR region are relatively weak and the resulting absorptions bands are an effect of a combination of different components present in the soil in variable proportions. Moreover, as stated above, both models were developed using different preprocessing techniques and a different number of factors, thus, direct comparison was difficult. Nevertheless, the visible part of the spectral range of both regression coefficients was dominated by the information originating from highly spectrally active iron oxide minerals at 530 and 620 nm, for clay and SSA, respectively (Scheinost, 1998). In both cases, absorption bands related to organic matter were present. For SSA, important wavelengths assigned to organic matter were present near 1650 and 1700 nm (assigned to aromatic C-H and aliphatic C-H groups, respectively) whereas, for clay near 825 and 1000 nm (assigned to aromatic C-H and amine N-H groups, respectively) (Stenberg et al., 2010). The presence of OM bands in both regression coefficients can be explained by the presence of the previously mentioned non-complexed SOC

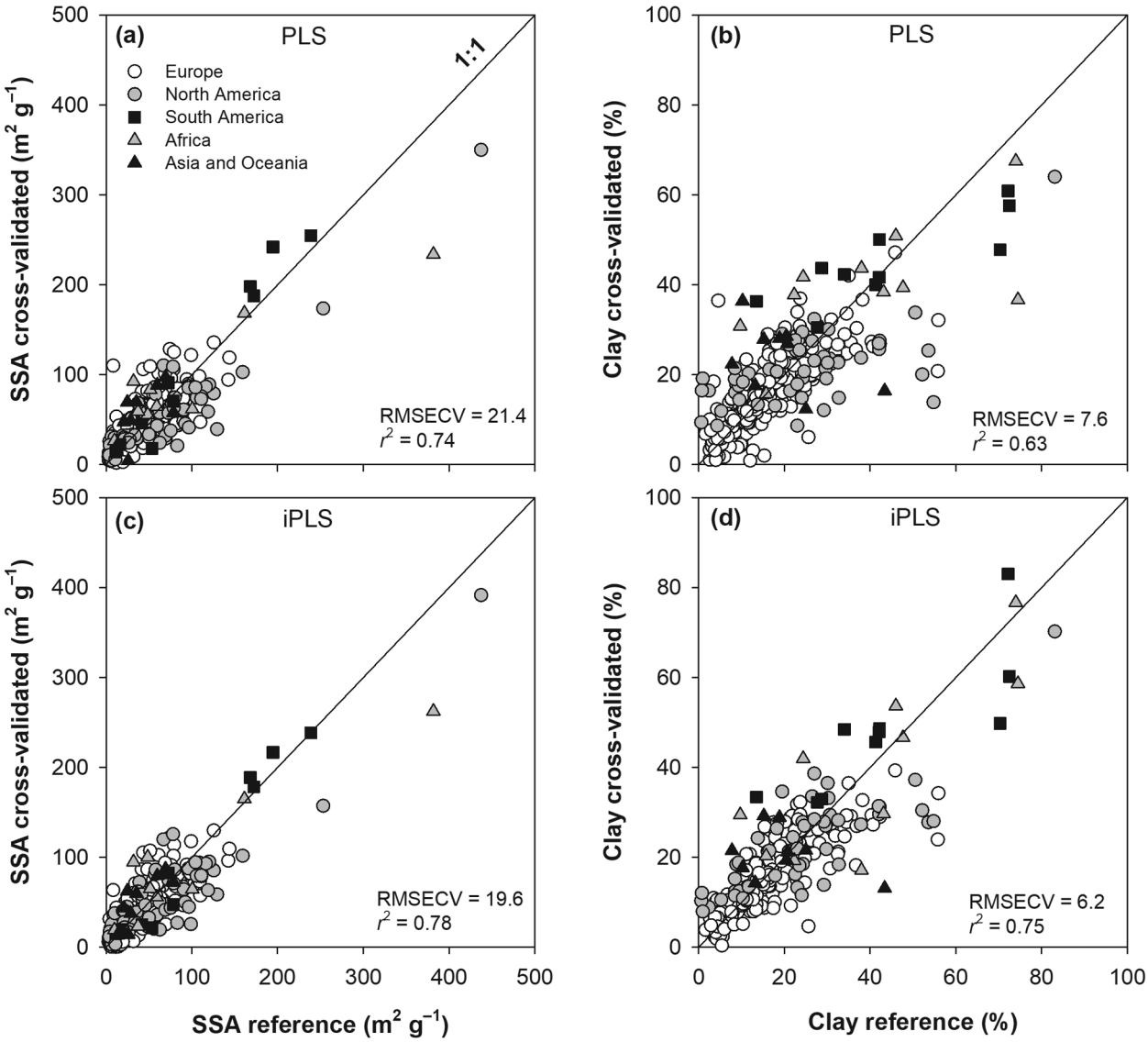

Fig. 4. Results of the best calibration (full cross-validation) models for soil specific surface area (SSA) and clay using Partial Least Squares Regression (PLS) (a and b, respectively) and interval Partial Least Squares Regression (iPLS) (c and d, respectively).

form in the majority of the European soil samples. Soil organic carbon in a non-complexed form can sorb onto the mineral phases covering and masking a large portion of the SSA of the mineral matrix (Kaiser and Guggenberger, 2003; Bartoli et al., 2007) resulting in a response from OM and not minerals in the vis – NIR spectrum. The presence of both negative and positive values of regression coefficients related to OM can be explained, as in the case of multiple linear regression for *n* ratio, by the fact that it is not only the quantity of OM but rather the different relationships between the quantity and quality of OM and clay. Thus, depending on the type of a functional group both positive and negative correlation with SSA and clay can be observed. Around 2200 nm an important peak related to the combination vibrations of O–H stretch and metal–OH bends (Stenberg et al., 2010) in the regression coefficient of SSA model was also present. Likewise, hydroxyl bonds were found to be important in the clay model showing an elevated value of regression coefficient near 1970 nm, which is indicative of smectites (Post and Noble, 1993; Clark, 1999).

Spectral Intervals for iPLS Calibration Models

As reported above, the calibration models for SSA and clay developed using the iPLS method were based on 50 nm intervals, accounting for 700 variables. More intervals (150 nm) and spectral variables were used in the iPLS clay model and accounted for a total of 1050 variables. The specific spectral ranges corresponding to the selected intervals are listed in Table 4. The identified intervals for the two parameters were assigned to similar yet not identical spectrally active components, which was also reported by Ben-Dor and Banin (1995). Furthermore, it appears that a higher number of smaller intervals was used for SSA estimation, whereas the intervals selected in clay modeling included a higher number of neighboring wavelengths indicating a higher specificity of SSA response and broader spectral regions related to clay, respectively. In general, the same type of the important soil components for the prediction of SSA and clay, as indicated by regression coefficients in the PLS models was found. In both cases, the important spectral intervals in the visible range of the spectrum were assigned to iron oxides. As expected, multiple

Table 3. Partial Least Squares regression (PLS) and interval PLS (iPLS) cross-validation results for clay content. †‡

Pretreatment	Method	IN	NV	RMSECV	R^2	Bias	NF
		nm		%			
No pretreatment (Abs.)	PLS		2100	9.0	0.48	–0.02	4
	iPLS	150	900	8.1	0.58	–0.03	6
		100	100	8.7	0.52	–0.02	5
		50	150	8.3	0.57	0.09	6
		25	75	8.7	0.52	–0.01	6
SNV	PLS		2100	8.5	0.51	0.001	6
	iPLS	150	450	8.2	0.53	–0.06	6
		100	700	6.7	0.69	–0.13	10
		50	300	7.2	0.53	–0.06	8
		25	275	6.6	0.71	–0.09	10
MSC	PLS		2100	8.6	0.53	0.01	6
	iPLS	150	750	7.9	0.60	–0.06	6
		100	700	6.7	0.72	–0.10	10
		50	200	7.9	0.61	–0.10	9
		25	75	7.5	0.64	0.01	9
first der.	PLS		2100	7.9	0.61	–0.05	6
	iPLS	150	150	8.2	0.58	–0.02	8
		100	400	8.5	0.55	–0.09	7
		50	100	7.7	0.63	–0.11	10
		25	75	7.4	0.65	–0.07	10
second der.	PLS		*2100*	*7.6*	*0.63*	*–0.05*	*7*
	iPLS	*150*	*1050*	*6.2*	*0.75*	*–0.09*	*10*
		100	1200	6.4	0.74	–0.04	10
		50	250	7.4	0.65	–0.13	9
		25	175	6.9	0.7	–0.02	10

† Abs., absorbance; first der. and second der., the first and second Savitzky-Golay derivatives; IN, number of intervals; iPLS- interval Partial Least Squares regression; MSC, multiplicative scatter correction; NF, number of factors; NV, number of variables; PLS, Partial Least Squares regression; RMSECV, Root Mean Square Error of Cross-Validation; SNV, standard normal variate.

‡ Numbers in italics denote the best models.

intervals covering wavelengths related to a combination of OH forms in water and in the mineral crystal lattice of smectites, illites, and kaolinites were selected for SSA and clay iPLS models in the NIR region, and were also listed by Ben-Dor and Banin (1995). However, most of these intervals could also be assigned to bands related to OM and because they were relatively broad it was not possible to identify if it was OH or OM, or the combination of the two components which was involved. No important wavelengths related to iron oxides for SSA and clay PLS models were found in the previous study by Ben-Dor and Banin (1995). This could be explained by the fact that no visible part of the spectrum was utilized in their investigation. They did not assign the important wavelengths to OM wavelengths either. However, the samples investigated in their study were more homogeneous in terms of the geographic location and thus might not have represented both forms of complexed and non-complexed SOC.

Independent Validation

The results of independent validation performed on 110 samples for SSA and clay determination are presented in Fig. 6a and 6b, respectively. Acceptable predictive ability for SSA for the independent validation data set using PLS was obtained (RMSEP = 14.2 $m^2\ g^{-1}$, R^2 = 0.76, and RPIQ = 1.8). Despite better performance of the iPLS calibration model for SSA than of the PLS model (Table 2), the independent validation using iPLS did not outperform significantly that of PLS (Fig. 6a). Similar RMSEP and R^2 values were obtained (13.6 $m^2\ g^{-1}$; 0.76). Likewise, acceptable validation for clay with PLS method was obtained (RMSEP = 5.1%, R^2 = 0.73). Also for clay lower prediction error was a result of a lower range in clay values in the validation data set (3 −50%) than in the calibration data set (1–83%). Application of the iPLS method for clay determination led to slightly better predictive ability (RMSEP = 4.6%; R^2 = 0.77) compared to the PLS prediction (Fig. 6b). The comparison of the SSA and clay models is not straightforward, as the modeled properties have different units and ranges. Nevertheless, the comparison of the RPIQ values indicated no significant differences among the models. The PLS models for SSA and clay generated RPIQ values of 1.7 and 1.6, respectively, whereas the iPLS models a value of 1.8 for SSA and 1.7 for clay.

Additionally, the performance of a simple linear regression model based on clay content of the calibration samples (SSA = −0.716 + 2.633 × clay) was tested using the same validation set. The performance of the regression model was quite good (RMSEP = 12.7 $m^2\ g^{-1}$ and R^2 = 0.78). Furthermore, the SSA values obtained from the clay-based linear regression model were statistically similar to those obtained from Vis–NIRS (Student's t test; p = 0.441). Nevertheless, the use of vis-NIRS is advanta-

geous as multiple soil properties can be determined including both basic and also more complex soil properties from one vis – NIR spectrum.

Ben-Dor et al. (1997) obtained a lower predictive ability (SEP = 50.2 $m^2\ g^{-1}$ and R^2 = 0.70) for a similar range of SSA values (11–425 $m^2\ g^{-1}$) on 12 groups of Israeli soils. However, the sensor used in their investigation covered only the NIR region (1000–2500nm). In another study on the in situ soil investigation performed with Vis–NIRS sensor on four profiles in Israel, higher predictive ability (RMSEP = 20.9 $m^2 g^{-1}$, R^2 = 0.92) for also a high range of SSA values (27–407 $m^2\ g^{-1}$) was achieved (Ben-Dor et al., 2008). Despite relatively similar SSA range in these and the present study, direct comparison is very difficult. The mentioned investigations used sensors with different spectral range, applied spectral measurements to differently prepared soils and smaller datasets, included subsoils, and were focused on relatively homogenous (in geographical sense) areas in comparison to our work.

CONCLUSIONS

In the present study, a large data set of soil samples originating from different parts of the world was investigated for testing the feasibility of using Vis–NIRS for SSA determination. Considering the extremely large range of SSA in the investigated soil samples acceptable predictive ability for the PLS independent validation was achieved (RPIQ = 1.8) and was slightly higher than for the independent clay validation (RPIQ = 1.6). The application of iPLS as a variable selection method resulted in little improved accuracy for both SSA and clay determinations (RPIQ of 1.8 and 1.7, respectively) in comparison to using PLS method and the entire vis–NIR spectrum. However, it decreased the number of variables significantly from 2100 to 700, and to 1050 for the determination of SSA and clay, respectively, thus, diminishing the complexity of the models. Furthermore, a higher number of shorter spectral intervals for SSA predictions was selected and indicated a higher specificity of SSA than of clay response in the vis-NIR spectral range. The different degree of saturation of clay minerals by organic matter in the investigated

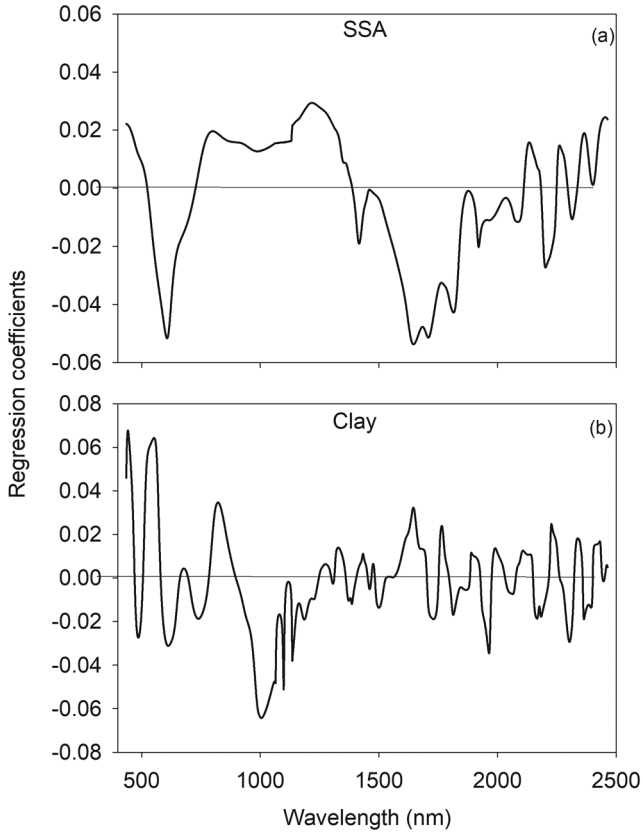

Fig. 5. Regression coefficients for Partial Least Squares calibration models of: a) soil specific surface area (SSA) (based on first derivative, 8 factors) and b) clay content (based on second derivative and 7 factors).

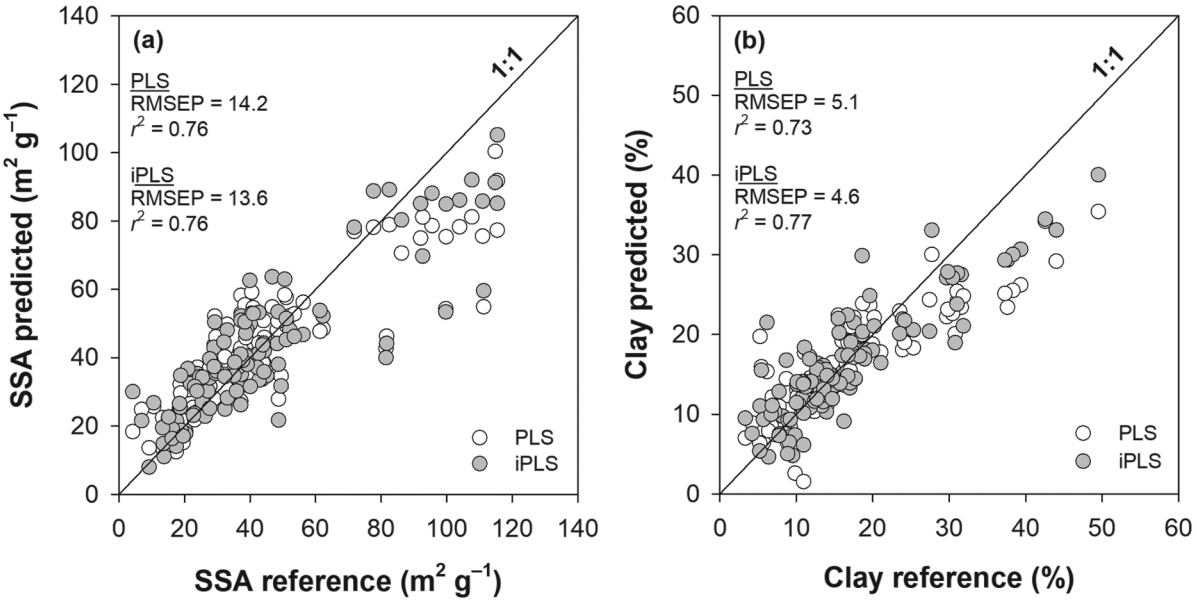

Fig. 6. Results of the independent validation for soil specific surface area (SSA) determinations and clay (a and b, respectively) using Partial Least Squares Regression (PLS) and interval Partial Least Squares Regression (iPLS).

Table 4. List of the important spectral intervals for the best SSA and clay iPLS models.†

	Spectral interval	Possible component	Reference
	nm		
SSA	750–900	Fe_{ox}	Scheinost, 1998
	750–900	Fe_{ox}/OM	Hunt, 1977; Stenberg et al., 2010
	1300–1350	OM	Ben-Dor et al., 1997; Ertlen et al., 2010
	1550–1650	OH/OM	Stenberg et al., 2010; Ben-Dor et al., 1997
	1850–1900	OH	Hunt, 1977; Stenberg et al., 2010
	2050–2100	OH/OM	Ben-Dor et al., 1997; Stenberg et al., 2010); Shonk et al., 1991
	2150–2250	OH/OM	Clark et al., 1990; Post and Nobel, 1993; Ben-Dor, 2002
	2350–2400	OH/OM	Ben-Dor et al., 1997; Stenberg et al., 2010; Post and Nobel, 1993
	2450–2500	OH	Clark, 1999; Stenberg et al., 2010; Post and Nobel, 1993
Clay	400–550	Fe_{ox}	Scheinost, 1998
	1150–1450	OM/OH	Shonk et al., 1991; Ertlen et al., 2010; Ben-Dor et al., 1997
	1600–1900	OH/OM	Stenberg et al., 2010; Ben-Dor et al., 1997; Hunt, 1977
	2200–2500	OH/OM	Hunt, 1977; Ben-Dor et al., 1997; Stenberg et al., 2010; Clark et al., 1990

† Fe_{ox}, iron oxides; OM, organic matter, SSA, specific surface area.

soils clearly induced changes in the SSA and was reflected in the spectral response of the organo-mineral composition and its interactions. Therefore, the important wavelengths in both PLS regression coefficients and the selected iPLS spectral intervals were not only assigned to Fe oxides and water bands related to different types of clay minerals, but also to organic matter, as a result of masking effect of the noncomplexed organic carbon on the mineral phases.

REFERENCES

Arthur, E., M. Tuller, P. Moldrup, M.H. Greve, M. Knadel, and L.W. de Jonge. 2018. Applicability of the GAB water vapour sorption model for estimation of soil specific surface area. Eur. J. Soil Sci. doi:10.1111/ejss.12524

Arthur, E., M. Tuller, P. Moldrup, A.C. Resurreccion, M.S. Meding, K. Kawamoto, T. Komatsu, and L.W. de Jonge. 2013. Soil specific surface area and non-singularity of soil-water retention at low saturations. Soil Sci. Soc. Am. J. 77:43–53. doi:10.2136/sssaj2012.0262

Barnes, E.M., K.A. Sudduth, J.W. Hummel, S.M. Lesch, D.L. Corwin, C. Yang, C.S.T. Daughtry, and W.C. Baush. 2003. Remote-and ground-based sensor techniques to map soil properties. Photogramm. Eng. Remote Sensing 69:619–630. doi:10.14358/PERS.69.6.619

Barnes, R.J., Dhanoa, M.S., and S.J. Lister. 1989. Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra. Appl. Spectrosc. 43:772–777. doi:10.1366/0003702894202201

Bartoli, F., A.J. Poulenard, and B.E. Schouller. 2007. Influence of allophane and organic matter contents on surface properties of Andosols. Eur. J. Soil Sci. 58:450–464. doi:10.1111/j.1365-2389.2007.00899.x

Bellon-Maurel, V., E. Fernandez-Ahumada, B. Palagos, J.-M. Roger, and A. McBratney. 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. TrAC, Trends Anal. Chem. 29:1073–1081. doi:10.1016/j.trac.2010.05.006

Ben-Dor, E., and A. Banin. 1995. Near-infrared analysis as a rapid method to simultaneusly evaluate several soil properties. Soil Sci. Soc. Am. J. 59:364–372. doi:10.2136/sssaj1995.03615995005900020014x

Ben-Dor, E., Y. Inbar, and Y. Chen. 1997. The reflectance spectra of organic matter in the visible near infrared and short wave infrared region (400-2500 nm) during a control decomposition process. Remote Sens. Environ. 61:1–15. doi:10.1016/S0034-4257(96)00120-4

Ben-Dor, E., D. Heller, and A. Chudnovsky. 2008. A novel method of classifying soil profiles in the field using optical means. Soil Sci. Soc. Am. J. 72:1113–1123. doi:10.2136/sssaj2006.0059

Ben-Dor, E. 2002. Quantitative remote sensing of soil properties. Adv. Agron. 75:173–243. doi:10.1016/S0065-2113(02)75005-0

Bishop, J.L., C.M. Pieters, and J.O. Edwards. 1994. Infrared spectroscopic analyses on the nature of water in montmorillonite. Adv. Agron. 42:702–716. doi: 10.1346/CCMN.1994.0420606

Borkovec, M., Wu, Q., Degovics, G., Laggner, P., Sticher, H. 1993. Surface-area and size distributions of soil particles. Colloids Surf., A 73: 65-76, doi: 10.1016/0927-7757(93)80007-2

Cerato, A., and A. Lutenegger. 2002. Determination of surface area of fine-grained soils by the Ethylene Glycol Monoethyl Ether (EGME) Method. Geotech. Test. J. 25:315–321. doi:10.1520/GTJ11087J

Clark, R.N. 1999. Spectroscopy of rocks and minerals and principles of spectroscopy. In: A.N. Rencz, editor, Remote sensing for the earth sciences. Manual of remote sensing. John Wiley and Sons, Chichester, UK. p. 3–58.

Clark, R.N., T.V.V. King, M. Klejwa, and A. Swayze. 1990. High spectral resolution reflectance spectrosocpy of minerals. J. Geophys. Res. 95:12653–12680. doi:10.1029/JB095iB08p12653

Dexter, A.R., G. Richard, D. Arrouays, E.A. Czyz, C. Jolivet, and O. Duval. 2008. Complexed organicmattercontrolssoilphysicalproperties Geoderma 144:620–627. doi:10.1016/j.geoderma.2008.01.022

Ding, F., J.G. Cai, M.S. Song, and P. Yuan. 2013. The relationship between organic matter and specific surface area in < 2 μm clay size fraction of muddy source rock. Sci. China Earth Sci. 56:1343-1349. doi:10.1007/s11430-013-4606-5

Ertlen, D., D. Schwartz, M. Trautmann, R. Webster, and D. Brunet. 2010. Discriminating between organic matter in soil from grass and forest by near-infrared spectroscopy. Eur. J. Soil Sci. 61:207–216. doi:10.1111/j.1365-2389.2009.01219.x

Gee, G.W., and D. Or. 2002. Particle-size analysis. In: J.H. Dane and G.C. Topp, editors, Methods of soil analysis. Part 4. Soil Science Society of America, Madison, WI. p. 255–293. doi:10.2136/sssabookser5.4.c12.

Gowen, A.A., G. Downey, C. Esquerre, and C.P. O'Donnell. 2011. Preventing over-fitting in PLS calibration models of near-infrared (NIR) spectroscopy data using regression coefficients. J. Chemometr. 25:375–381. doi:10.1002/cem.1349

Hang, P.T. and Brindley, G.W. 1970. Methylene blue absorption by clay minerals. Determination of surface areas and cation exchange capacities (Clay-organic studies XVIII). Clays Clay Miner. 18: 203-212.

Heister, K. 2014. The measurement of the specific surface area of soils by gas and polar liquid adsorption methods-limitation and potentials. Geoderma 216:75–87. doi:10.1016/j.geoderma.2013.10.015

Hermansen, C., M. Knadel, P. Moldrup, M.H. Greve, R. Gislum, and L.W. de Jonge. 2016. Visible–near-infrared spectroscopy can predict the clay/organic carbon and mineral fines/organic carbon ratios. Soil Sci. Soc. Am. J. 80:1486–1495. doi:10.2136/sssaj2016.05.0159

Hermansen, C., M. Knadel, P. Moldrup, M.H. Greve, D. Karup, and L.W. de Jonge. 2017. Complete soil texture is accurately predicted by visible near-infrared spectroscopy. Soil Sci. Soc. Am. J. 81. doi:10.2136/sssaj2017.02.0066

Hunt, G. 1977. Spectral signatures of particulate minerals in the visible and near infrared. Geophysics 42:501–513. doi:10.1190/1.1440721

Kaiser, K., and G. Guggenberger. 2003. Mineral surfaces and soil organic matter. Eur. J. Soil Sci. 54:219–236. doi:10.1046/j.1365-2389.2003.00544.x

Kaiser, K., and W. Zech. 1998. Soil dissolved organic matter sorption as influenced

by organic and sesquioxide coatings and sorbed sulfate. Soil Sci. Soc. Am. J. 62:129–136. doi:10.2136/sssaj1998.03615995006200010017x

Katuwal, S., C. Hermansen, M. Knadel, P. Moldrup, M.H. Greve, and L.W. de Jonge. 2017. Combining X-ray computed tomography and visible near-infrared spectroscopy for prediction of soil structural properties. Vadose Zone J. doi:10.2136/vzj2016.06.0054

Kennard, R.W., and L.A. Stone. 1969. Computer aided design of experiments. Technometrics 11:137–148. doi:10.1080/00401706.1969.10490666

Kim, K.C., Yoon, T., Bae, Y. 2016. Applicability of using CO_2 adsorption isotherms to determine BET surface areas of microporous materials. Microporous Mesoporous Mater. 224:294-301, doi:10.1016/j.micromeso.2016.01.003

Kim, I., R.R. Pullanagari, M. Deurer, R. Singh, K.Y. Huh, and B.E. Clothier. 2014. The use of visible and near-infrared spectroscopy for the analysis of soil water repellency. Eur. J. Soil Sci. 65:360–368. doi:10.1111/ejss.12138

Knadel, M., F. Masís-Meléndez, L. Wollesen de Jonge, P. Moldrup, E. Arthur, and M. Humlekrog Greve. 2016. Assessing soil water repellency of a sandy field with visible near infrared spectroscopy. J. Near Infrared Spectrosc. 24:215–224. doi:10.1255/jnirs.1188

Martens, H., and T. Næs. 1989. Multivariate calibration. John Wiley and Sons, New York. doi:10.1002/0471667196.ess1105.

Moberg, J.P. 1990. Composition and development of the clay fraction in Danish soils. An overview. Sci. Geol. Bull. 43:193–202.

Nelson, D.W., and L.E. Sommers. 1996. Total carbon, organic carbon, and organic matter. In: D.L. Sparks, A.L. Page, P.A. Helmke, and R.H. Loeppert, editors, Methods of soil analysis part 3– Chemical cethods. Soil Science Society of America, Madison, WI. p. 961–1010.

Nørgaard, L., A. Saudland, J. Wagner, J.P. Nielsen, L. Munck, and S.B. Engelsen. 2000. Interval Partial Least-Squares Regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. Appl. Spectrosc. 54:413–419. doi:10.1366/0003702001949500

Paradelo, M., C. Hermansen, M. Knadel, P. Moldrup, M. Greve, and L. de Jonge. 2016. Field-scale predictions of soil contaminant sorption using visible–near infrared spectroscopy. J. Near Infrared Spectrosc. 24:281–291. doi:10.1255/jnirs.1228

Pennell, K.D. 2002. Specific surface area. In: J.H. Dane and G.C. Topp, editors, Methods of soil analysis, Part 4: Physical methods. Soil Science Society of America, Madison, WI. p. 295–315. doi:10.1016/B978-0-12-409548-9.09583-X.

Petersen, L.W., P. Moldrup, O.H. Jacobsen, and D.E. Rolston. 1996. Relations between specific surface area and soil physical and chemical properties. Soil Sci. 161:9–21. doi:10.1097/00010694-199601000-00003

Post, J.L., and P.N. Noble. 1993. The near-infrared combination band frequencies of dioctahedral smectites, micas, and illites. Clays Clay Miner. 41:639–644. doi:10.1346/CCMN.1993.0410601

Quirk, J.P., and R.S. Murray. 1999. Appraisal of the ethylene glycol monoethyl ether method for measuring hydratable surface area of clays and soils. Soil Sci. Soc. Am. J. 63:839–849. doi:10.2136/sssaj1999.634839x

Sadeghi, M., E. Babaeian, M. Tuller, and S.B. Jones. 2018. Particle size effects on soil reflectance explained by an analytical radiative transfer model. Remote Sens. Environ. 210:375–386. doi:10.1016/j.rse.2018.03.028

Savitzky, A., and M. Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36:1627–1639. doi:10.1021/ac60214a047

Scheinost, A.C. 1998. Use and limitations of second-derivative diffuse reflectance spectroscopy in the visible to near-infrared range to identify and quantify Fe oxide minerals in soils. Clays Clay Miner. 46: 528-536, doi:10.1346/CCMN.1998.0460506

Shonk, G.A., Gaultney, L.D., Schulze, D.G. and Van Scoyoc, G.E. 1991. Spectroscopic sensing of soil organic matter content. Trans. ASAE 34:1978–1984. doi:10.13031/2013.31826</jrn>

Soriano-Disla, J.M., L.J. Janik, R.A. Viscarra Rossel, L.M. Macdonald, and M.J. McLaughlin. 2013. The Performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. Appl. Spectrosc. Rev. 49:139–186. doi:10.1080/05704928.2013.811081

Stenberg, B., R.A.V. Rossel, A.M. Mouazen, and J. Wetterlind. 2010. Visible and near infrared spectroscopy in soil science. Adv. Agron. 107:163-215, doi:10.1016/S0065-2113(10)07005-7

Tiller, K.G., and L.H. Smith. 1990. Limitations of EGME retention to estimate the surface area of soils. Aust. J. Soil Res. 28:1–26. doi:10.1071/SR9900001

Tuller, M., and D. Or. 2005. Water films and scaling of soil characteristic curves at low water contents. Water Resour. Res. 41. doi:10.1029/2005WR004142

Viscarra Rossel, R.A., and W.S. Hicks. 2015. Soil organic carbon and its fractions estimated by visible-near infrared transfer functions. Eur. J. Soil Sci. 66:438–450. doi:10.1111/ejss.12237

Webster, R. 2001. Statistics to support soil research and their presentation. Eur. J. Soil Sci. 52:331–340. doi:10.1046/j.1365-2389.2001.00383.x